Modelling the Marathon

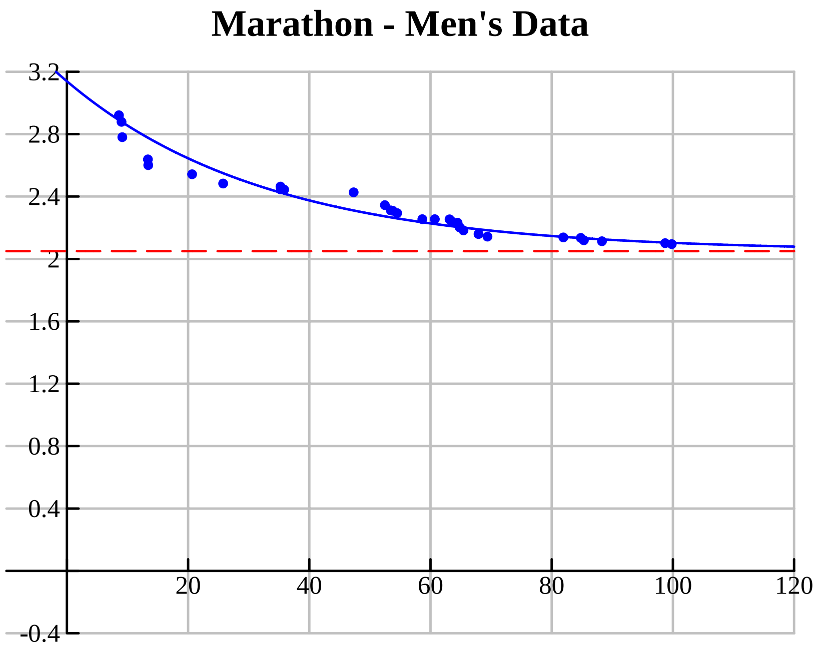
Men's Data

Marathon - Men's Data							
Date	Days since 1900	Years since 1900	Hours	Minutes	Seconds	Decimal hours	Modified
24/07/08	3128	8.5640	2	55	18	2.9217	.8717
1/01/09	3289	9.0048	2	52	45	2.8792	.8292
12/02/09	3331	9.1198	2	46	52	2.7811	.7311
12/05/13	4881	13.3634	2	38	16	2.6378	.5878
31/05/13	4900	13.4155	2	36	6	2.6017	.5517
22/08/20	7540	20.6434	2	32	35	2.5431	.4931
12/10/25	9417	25.7823	2	29	1	2.4836	.4336
31/03/35	12874	35.2471	2	27	49	2.4636	.4136
3/04/35	12877	35.2553	2	26	44	2.4456	.3956
3/11/35	13091	35.8412	2	26	42	2.4450	.3950
19/04/47	17276	47.2991	2	25	39	2.4275	.3775
14/06/52	19159	52.4545	2	20	42	2.3450	.2950
13/06/53	19523	53.4511	2	18	40	2.3111	.2611
4/10/53	19636	53.7604	2	18	34	2.3094	.2594
26/06/54	19901	54.4860	2	17	39	2.2942	.2442
24/08/58	21421	58.6475	2	15	17	2.2547	.2047
10/09/60	22169	60.6954	2	15	16	2.2544	.2044
17/02/63	23059	63.1321	2	15	15	2.2542	.2042
15/06/63	23177	63.4552	2	14	28	2.2411	.1911
13/06/64	23541	64.4517	2	13	55	2.2319	.1819
21/10/64	23671	64.8077	2	12	11	2.2031	.1531
12/06/65	23905	65.4483	2	11	0	2.1833	.1333
3/12/67	24809	67.9233	2	9	36	2.1600	.1100
30/05/69	25353	69.4127	2	8	34	2.1428	.0928
6/12/81	29926	81.9329	2	8	18	2.1383	.0883
21/10/84	30976	84.8077	2	8	5	2.1347	.0847
20/04/85	31157	85.3032	2	7	12	2.1200	.0700
17/04/88	32250	88.2957	2	6	50	2.1139	.0639
20/09/98	36058	98.7214	2	6	5	2.1014	.0514
24/10/99	36457	99.8138	2	5	42	2.0950	.0450

Men's Data: Years and Modified time
Exponential Regression
regEQ(x) = 1.08868*.970245^x
a = 1.08868
b = .970245
r = -.976346

 Flinders Christian Community College, Victoria

Plotting the results assuming an **asymptote of 2.05 hours.**



Changing the expression into the form $T = a \cdot e^{bt} + 2.05$

$\ln(b\_) \rightarrow b$

-.030207

Hence the rule is $T(x) := (1.08868) \cdot e^{(-0.030207 \cdot x)} + 2.05$ "Done"

Checking the rule:

Years	Time	Rule	% Difference
8.564	2.9217	2.8905	-1.08
9.0048	2.8792	2.8794	.01
9.1198	2.7811	2.8765	3.32
13.3634	2.6378	2.7771	5.02
13.4155	2.6017	2.7759	6.28
20.6434	2.5431	2.6336	3.43
25.7823	2.4836	2.5497	2.59
35.2471	2.4636	2.4254	-1.57
35.2553	2.4456	2.4253	-.84
35.8412	2.445	2.4187	-1.09
47.2991	2.4275	2.3109	-5.05
52.4545	2.345	2.2732	-3.16
53.4511	2.3111	2.2666	-1.96
53.7604	2.3094	2.2646	-1.98
54.486	2.2942	2.2599	-1.52
58.6475	2.2547	2.2351	-.87
60.6954	2.2544	2.2240	-1.37
63.1321	2.2542	2.2117	-1.92
63.4552	2.2411	2.2101	-1.40
64.4517	2.2319	2.2054	-1.20
64.8077	2.2031	2.2037	.03
65.4483	2.1833	2.2008	.79
67.9233	2.16	2.1899	1.37
69.4127	2.1428	2.1837	1.88
81.9329	2.1383	2.1416	.16
84.8077	2.1347	2.1340	-.03
85.3032	2.12	2.1328	.60
88.2957	2.1139	2.1256	.55
98.7214	2.1014	2.1052	.18
99.8138	2.095	2.1034	.40
		Average	.05

Note that positive percentage differences mean that the rule under estimates the time obtained whilst negative percentage differences over estimates the time obtained.

The overall average of the percentage difference indicates that the rule is a reasonable model for the data.

Women's Data: 1964 - 1985

The data used for this model covers the years from 1964 until 1985. A subsequent model will include the most current data.

Marathon - Women's Data	1964 - 1985						
Date	Days since 1900	Years since 1900	Hours	Minutes	Seconds	Decimal hours	Modified
23/05/64	23520	64.3943	3	27	45	3.4625	1.4125
21/07/64	23579	64.5558	3	19	33	3.3258	1.2758
6/05/67	24598	67.3457	3	15	22	3.2561	1.2061
16/07/67	24669	67.5400	3	7	26	3.1239	1.0739
28/06/70	25747	70.4914	3	2	53	3.0481	.9981
9/05/71	26062	71.3539	3	1	42	3.0283	.9783
31/08/71	26176	71.6660	2	46	30	2.7750	.7250
27/10/74	27329	74.8227	2	46	24	2.7733	.7233
1/12/74	27364	74.9185	2	43	54	2.7317	.6817
21/04/75	27505	75.3046	2	42	42	2.7117	.6617
3/05/75	27517	75.3374	2	40	15	2.6708	.6208
12/10/75	27679	75.7810	2	38	19	2.6386	.5886
1/05/77	28246	77.3333	2	35	15	2.5875	.5375
10/09/77	28378	77.6947	2	34	47	2.5797	.5297
22/10/78	28785	78.8090	2	32	29	2.5414	.4914
21/10/79	29149	79.8056	2	27	32	2.4589	.4089
26/10/80	29520	80.8214	2	25	41	2.4281	.3781
28/10/81	29887	81.8261	2	25	29	2.4247	.3747
17/04/83	30423	83.2936	2	25	29	2.4247	.3747
18/04/83	30424	83.2964	2	22	42	2.3783	.3283
21/04/85	31158	85.3060	2	21	6	2.3517	.3017

Marathon: Women's Data 1964 - 1985

Exponential Regression

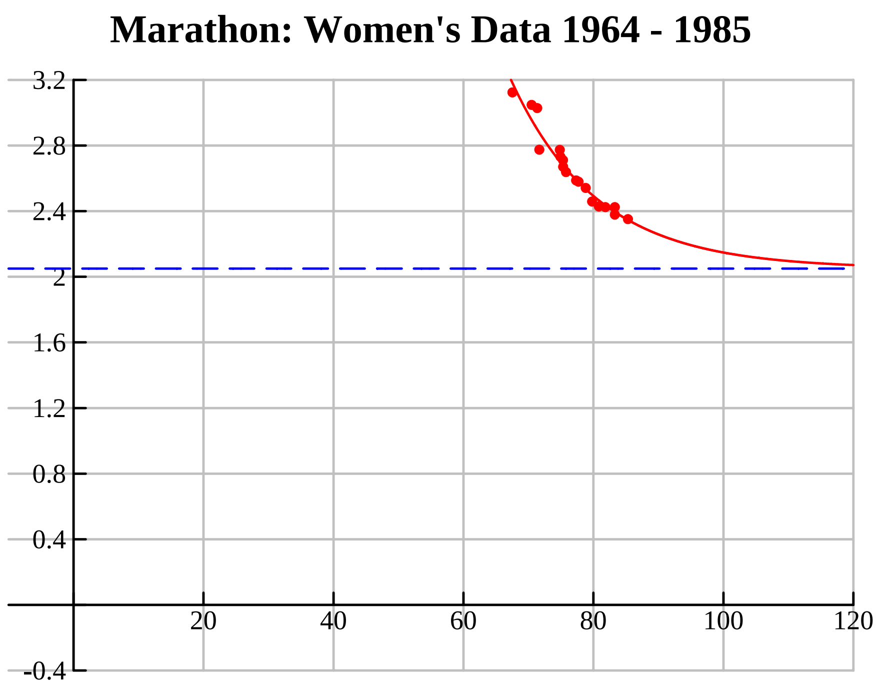
regEQ(x) = 182.956*.92745^x

a = 182.956

b = .92745

r = -.98753

Plotting the results assuming an **asytmptote of 2.05 hours.**



Changing the expression into the form $T = a \cdot e^{bt} + 2.05$

$\ln(b\_) \rightarrow b$

-.075316

Hence the rule is $T(x) := (182.956) \cdot e^{(-0.075316 \cdot x)} + 2.05$ "Done"

 Flinders Christian Community College, Victoria

Checking the rule:

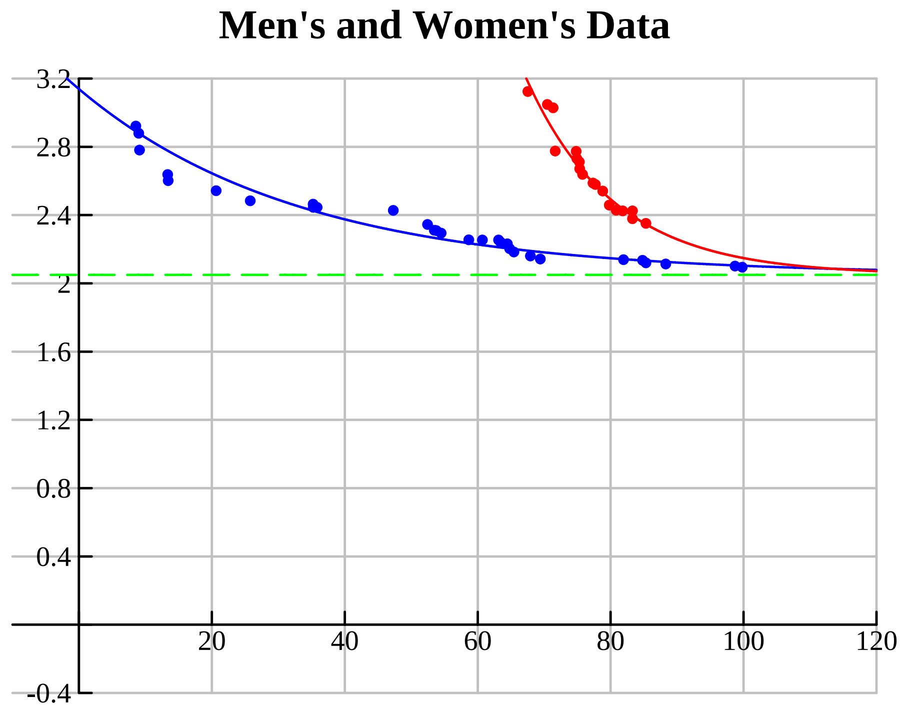
Years	Time	Rule	% Difference
64.3943	3.4625	3.4824	.57
64.5558	3.3258	3.4650	4.02
67.3457	3.2561	3.1969	-1.85
67.54	3.1239	3.1802	1.77
70.4914	3.0481	2.9549	-3.15
71.3539	3.0283	2.8980	-4.50
71.666	2.775	2.8783	3.59
74.8227	2.7733	2.7031	-2.60
74.9185	2.7317	2.6984	-1.24
75.3046	2.7117	2.6798	-1.19
75.3374	2.6708	2.6782	.28
75.781	2.6386	2.6576	.71
77.3333	2.5875	2.5905	.12
77.6947	2.5797	2.5760	-.14
78.809	2.5414	2.5337	-.30
79.8056	2.4589	2.4987	1.59
80.8214	2.4281	2.4657	1.52
81.8261	2.4247	2.4354	.44
83.2936	2.4247	2.3950	-1.24
83.2964	2.3783	2.3950	.70
85.306	2.3517	2.3465	-.22
		Average	-.05

Once again the model is fairly accurate.

Tabulating the results of both the Men's and Women's marathon and then plotting them on the same set of axes:

Men's Data			**Women's Data**		
Days since 1900	**Years since 1900**	**Decimal hours**	**Days since 1900**	**Years since 1900**	**Decimal hours**
3128	8.564	2.9217	23520	64.3943	3.4625
3289	9.0048	2.8792	23579	64.5558	3.3258
3331	9.1198	2.7811	24598	67.3457	3.2561
4881	13.3634	2.6378	24669	67.5400	3.1239
4900	13.4155	2.6017	25747	70.4914	3.0481
7540	20.6434	2.5431	26062	71.3539	3.0283
9417	25.7823	2.4836	26176	71.6660	2.7750
12874	35.2471	2.4636	27329	74.8227	2.7733
12877	35.2553	2.4456	27364	74.9185	2.7317
13091	35.8412	2.4450	27505	75.3046	2.7117
17276	47.2991	2.4275	27517	75.3374	2.6708
19159	52.4545	2.3450	27679	75.7810	2.6386
19523	53.4511	2.3111	28246	77.3333	2.5875
19636	53.7604	2.3094	28378	77.6947	2.5797
19901	54.486	2.2942	28785	78.8090	2.5414
21421	58.6475	2.2547	29149	79.8056	2.4589
22169	60.6954	2.2544	29520	80.8214	2.4281
23059	63.1321	2.2542	29887	81.8261	2.4247
23177	63.4552	2.2411	30423	83.2936	2.4247
23541	64.4517	2.2319	30424	83.2964	2.3783
23671	64.8077	2.2031	31158	85.3060	2.3517
23905	65.4483	2.1833			
24809	67.9233	2.1600			
25353	69.4127	2.1428			
29926	81.9329	2.1383			
30976	84.8077	2.1347			
31157	85.3032	2.1200			
32250	88.2957	2.1139			
36058	98.7214	2.1014			
36457	99.8138	2.0950			

Plotting the results assuming an **asymptote of 2.05 hours.**



In order to determine whether the Women's time will be faster than the Men's time we can calculate the point of intersection of the two functions.

The point of intersection is (113.598, 2.08521)

This corresponds to 2013 in the month of July.

However, using the last three records since 1985 in the model changes the results considerably:

Firstly, using the Women's rule to predict the times taken in the years from 1985 to 2001.

As we have the actual data for these years we can check whether what we thought would happen actually did or did not happen.

The result for the years 1985 to 2001 are as follows:

Date	Days since 1900	Years since 1900	Hours	Minutes	Seconds	Decimal hours
19/04/98	35904	98.2998	2	20	5	2.3347
26/09/99	36429	99.7372	2	20	4	2.3344
30/09/01	37164	101.7495	2	19	5	2.3181

The rule for the Women's data is $T(x) := (182.956) \cdot e^{(-0.075316 \cdot x)} + 2.05$ "Done"

Substituting the value from 1998 into the rule:

$T(98.2998)$ 2.16144

To convert to hours minutes and seconds:

$\text{fpart}(\text{ans}) \cdot 60$ 9.68618

$\text{fpart}(\text{ans}) \cdot 60$ 41.1707

Hence the prediction is 2 hours 9 minutes and 41 seconds.

This is significantly faster than the actual time of 2 hours 20 minutes and 5 seconds.

Substituting the values for 1999 and 2001 yield a similar result.

$T(99.7372)$ 2.15

$T(101.7495)$ 2.13594

Consequently it appears that the Women's record has not come down at the rate predicted by the original rule. As such, it appears that the initial prediction of 2013 may be premature.

Generating a rule using all of the current data:

Marathon - Men's Data			Marathon - Women's Data		
Days since 1900	Years since 1900	Decimal hours	Days since 1900	Years since 1900	Decimal hours
3128	8.5640	2.9217	23520	64.3943	3.4625
3289	9.0048	2.8792	23579	64.5558	3.3258
3331	9.1198	2.7811	24598	67.3457	3.2561
4881	13.3634	2.6378	24669	67.54	3.1239
4900	13.4155	2.6017	25747	70.4914	3.0481
7540	20.6434	2.5431	26062	71.3539	3.0283
9417	25.7823	2.4836	26176	71.666	2.775
12874	35.2471	2.4636	27329	74.8227	2.7733
12877	35.2553	2.4456	27364	74.9185	2.7317
13091	35.8412	2.4450	27505	75.3046	2.7117
17276	47.2991	2.4275	27517	75.3374	2.6708
19159	52.4545	2.3450	27679	75.781	2.6386
19523	53.4511	2.3111	28246	77.3333	2.5875
19636	53.7604	2.3094	28378	77.6947	2.5797
19901	54.4860	2.2942	28785	78.809	2.5414
21421	58.6475	2.2547	29149	79.8056	2.4589
22169	60.6954	2.2544	29520	80.8214	2.4281
23059	63.1321	2.2542	29887	81.8261	2.4247
23177	63.4552	2.2411	30423	83.2936	2.4247
23541	64.4517	2.2319	30424	83.2964	2.3783
23671	64.8077	2.2031	31158	85.306	2.3517
23905	65.4483	2.1833	35904	98.2998	2.3347
24809	67.9233	2.1600	36429	99.7372	2.3344
25353	69.4127	2.1428	37164	101.7495	2.3181
29926	81.9329	2.1383			
30976	84.8077	2.1347			
31157	85.3032	2.1200			
32250	88.2957	2.1139			
36058	98.7214	2.1014			
36457	99.8138	2.0950			

Women's Data
Exponential Regression
regEQ(x) = 21.8794*.9545^x
a = 21.8794
b = .9545

Changing the expression into the form $T = a \cdot e^{bt} + 2.05$

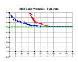
$\ln(b\_) \rightarrow b$

$-.046568$

Hence the rules are

Women's Data: $W(x) := (21.8794) \cdot e^{(-0.04568 \cdot x)} + 2.05$ "Done"

Men's Data: $M(x) := (1.08868) \cdot e^{(-0.030207 \cdot x)} + 2.05$ "Done"



In this case the point of intersection of the two functions is (183.399, 2.05428)

This suggests that the Women's times will be slower than the Men's times until around April 2083.

Solving the simultaneous equations:

Men's result: using the points (9.0027, 2.8792) and (65.4466, 2.1817)

$2.8792 = a \cdot e^{(9.0048 \cdot b)} + 2.05$ Equation 1

$2.1833 = a \cdot e^{(65.4483 \cdot b)} + 2.05$ Equation 2

$\text{solve}\left(2.8792 = a \cdot e^{(9.0048 \cdot b)} + 2.05 \text{ and } 2.1833 = a \cdot e^{(65.4483 \cdot b)} + 2.05, \{a, b\}\right)$

a = 1.10995 and b = -.032384
Warning: More solutions may exist
Warning: Overflow replaced by infinity or -infinity

Hence the rule for the Men's data is:

$TM(x) := 1.10995 \cdot e^{(-0.032384 \cdot x)} + 2.05$ "Done"

Women's result: using the points (64.5558, 3.3258) and (80.8214, 2.4281)

$3.3258 = a \cdot e^{(64.5558 \cdot b)} + 2.05$ Equation 1

$2.4281 = a \cdot e^{(80.8214 \cdot b)} + 2.05$ Equation 2

$\text{solve}\left(3.3258 = a \cdot e^{(64.5558 \cdot b)} + 2.05 \text{ and } 2.4281 = a \cdot e^{(80.8214 \cdot b)} + 2.05, \{a, b\}\right)$

a = 159.234 and b = -.074769
Warning: More solutions may exist

Hence the rule for the Women's data is:

$TW(x) := 159.234 \cdot e^{(-0.074769 \cdot x)} + 2.05$ "Done"

Finding the point of intersection of these two functions:

$\text{nsolve}(TW(x) = TM(x), x)$

117.166

This suggests that, using the given data points, the Women's and Men's marathon times will be equal in around 2017. This is reasonably close to the first estimation based on the regression model of the Men's times against the Women's times to 1985.
Some of the limitations of using this approach are:

a) The data points selected will influence the result quite significantly.

b) With exponential models there can be significant variation in the long-term result despite only minor changes in the short-term.

c) We are also assuming that the asymptote for the data is 2.05 hours. There was no formalised method for determining this number, rather an estimate of a reasonable time was made. This could be analogous to the "4 minute mile" barrier that was held for so long. However, once the first person broke this time a large number of others also broke it in the few years following.

d) We are not allowing for large "leaps forward" in performance caused by improved training regimes and improvements in timing equipment or sporting equipment. The assumption is that the trend will continue as it has done for years.

e) We are also assuming that the exponential model is the best model. While it appears to fit the data satisfactorily, there may be a better model.